EPA Response to Peer Review Comments received on EPA's Draft Assessment of Detection and Quantitation Limit Concepts

February 25, 2003

Introduction

In August 2002 EPA conducted a peer review of a draft "Technical Support Document of Detection and Quantitation Regulations under the Clean Water Act" (the "Assessment Document"). This peer review was conducted in accordance with EPA's peer review policies described in the Science Policy Council Handbook (EPA 100-B-00-001). The review was performed by two experts in the field of analytical chemistry and two experts in the statistical aspects of analytical data interpretation. Each reviewer was provided with a draft version of the Assessment Document, which documented the Agency's approach to the assessment and the Agency's preliminary findings and conclusions. Reviewers also were provided with copies of all data evaluated in the assessment, statistical programs used to analyze the data, and copies of the detection and quantitation concepts and procedures evaluated by EPA. In accordance with the Agency's peer review policies, the reviewers were provided with a written 'charge' intended to ensure the evaluation would meet EPA needs.

In its charge to the peer reviewers, EPA requested a written evaluation of whether the assessment approach described by EPA was valid and conceptually sound. Reviewers also were asked to consider and address eight specific questions pertaining to the adequacy of the concepts and issues considered, the evaluation criteria developed by EPA, EPA's assessment and conclusions, the data used to perform the assessment, suggested improvements to the procedures discussed, and EPA's consideration of interlaboratory vs. intralaboratory issues. Copies of all materials associated with the peer review, including the peer review charge, the materials provided to the peer reviewers for review, and complete copies of the peer reviewers' comments are provided separately in the public docket supporting the Agency's assessment.

Responses to Comments Submitted by David Rocke

EPA's reassessment is complete and sound (DR-1)

Comment: The list of detection and quantitation concepts, the issues laid out in Chapter 3 of the peer review version of the Assessment Document, and the evaluation criteria in Chapter 4 are sufficiently complete and adequate for the reassessment. The assessments in Chapter 5 are sound.

Response: EPA appreciates the support for the reassessment.

Purpose of detection limit (DR-2, DR-14)

Comment: The purpose of a detection limit in Criterion 4 of the peer review version of the Assessment Document appears from the description and the discussion to be a mix of the Currie concepts of critical value (CRV) and minimum detectable value (MDV). What should appear here is the CRV equivalent. Here is a suggested re-wording:
The detection level concept should identify the signal or estimated concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory.

Response: EPA agrees and has revised the purpose along the lines that the peer reviewer suggests.

Quantitation limits are misnomer (DR-3, DR-9, DR-15)

Comment: With respect to the limit of quantitation concept, the EPA ML is as good as any of the others given. However, all are flawed by the assumption that there is some level higher than the Currie critical value (CRV) needed before quantitative assessments can be made. This is not supported in this document, nor anywhere else I have seen, except as an almost unexamined assumption. After all, the concentration at which the MDL is measured must generate peaks that can be measured. If the instrument can be read, and the spectra can be recognized, then the ML is exceeded, regardless of the other issues. The entire concept of a quantitation level higher than the CRV should be immediately discarded.

Criterion 5 of the peer review version of the Assessment Document should be completely changed to reflect that almost all implementations of limits of quantitation have nothing to do with whether the measurements are actually quantitative. No arbitrary standard regarding multiples of the standard deviation at zero or a desired CRV is appropriate for any purpose in analytical chemistry or the regulation of toxic substances. This includes the PQL, the AML and other related methods. None of them generate a useful number. Here is a suggested re-wording:

The quantitation limit concept should identify a concentration at which the instrument yields a measurable signal at least 99% of the time, and which is also no smaller than the detection level. This will often be the same as the detection level.

Response: EPA agrees in principle but does not want to create a conflict with historical precedent and present use of a quantitation limit as understood and used by the analytical community. EPA agrees with the reviewer's comments that all results above the critical value (or MDL) are quantitative measurements. At issue is the potential degree of uncertainty associated with a specific value. While there are many who advocate reporting all results down to the MDL, the EPA Office of Water believes that in a regulatory context, the greater uncertainty associated with measurements at the MDL warrants the use of a "quantitation limit" as a means to assure that the uncertainty of a reported value is at an acceptable level. Therefore, EPA has not revised Criterion 5 and has not discarded the quantitation limit.

MDL is reasonable (DR-4, DR-16)

Comment: The MDL is a reasonable implementation of Currie's critical value (CRV) concept for situations in which the instrument may not yield reliable data for blanks. With a slight alteration to the specifications on the spike concentration, the EPA MDL as now given is a reasonable, practical implementation of a limit of detection concept and procedure. None of the other procedures is an improvement on this overall.

Response: EPA agrees and thanks the reviewer.

Additional data sets (DR-5)

Comment: There is no need to examine additional data sets.

Response: EPA believes that the data sets examined for the peer review are sufficient to allow the reassessment. However, EPA is soliciting further data sets in the proposal in the event that there could be more definitive data sets that may aid in further assessing detection and quantitation limit concepts.

Interlaboratory variability (DR-6, DR-12)

Comment: EPA's position on interlaboratory vs. intralaboratory variability is reasonable. If a laboratory computes a CRV using a procedure such as the MDL, it makes no sense to expand this to account for interlaboratory variability. Whether other labs can or cannot detect the substance with a signal at the MDL of the given laboratory is irrelevant. It may be different if the goal is precisely to determine the quantity of the analyte in a standard sample. In this case, interlaboratory variability may be appropriately considered. It should not be considered in detection decisions unless it can

be shown that such decisions in an individual laboratory are biased and may over- or under-estimate the true critical value (detection threshold) in that laboratory.

For the specific purpose of determining whether a given sample exceeds the safe level, a general interlaboratory study is not of much use, since it may be influenced by the performance of laboratories at levels far removed from the point at issue. If the safe level is below a critical value, use of the critical value is appropriate as an action threshold. If the safe level is above the critical value, interlaboratory variation should only be taken into account if it can be shown that the number of false positives when the analyte is present at the safe level is not well controlled using the usual intralaboratory calibration methods.

Response: EPA agrees.

Randomized study design (DR-7)

Comment: EPA's tests of variability vs concentration that resulted in the ICP/MS, Episode 6000, and 6184 data sets were conducted in sequence from high to low concentration. This grouped analysis by concentration leads to anomalous results. If all samples at a given concentration are analyzed in sequence, then the next concentration, and so on, the values at a given concentration will be closer together than would be the case if they were analyzed at different times, or interspersed with other concentrations. This problem should be fixed by using a proper randomized-order design, but can be mitigated by always looking at variability around the calibration line, rather than around the mean of the replicates.

Response: The reason concentrations were not randomized was because of a concern about cross contamination of low-concentration samples by high-concentration samples. For example, in Episode 6000 study, the concentrations ranged from 0.1 to 100 times the initial estimate of the MDL, or a concentration range of a factor of 1000. If a highest concentration sample were analyzed and followed by a lowest concentration sample, the possibility exists that some of the analyte would be carried into the lowest concentration sample, inflating the result, and increasing the apparent variability and the resulting detection and/or quantitation limit at the lowest concentration.

Regarding use of variability around the calibration line, the measured concentrations in the various variability vs concentration studies were determined using a calibration factor, response factor, or calibration curve, as appropriate to the analytical technique and as required by the analytical method. This means of determining the analyte concentration assured that the variability was mitigated, as the reviewer suggests.

In response to this reviewers comment, EPA has revised the Assessment Document to further explain this issue.

MDL procedure should allow use of blanks (DR-8, DR-16)

Comment: The MDL had a number of problems that needed repair, some of which were fixed in the rewording on page 5-4 of the peer review version of the Assessment Document. The basic concept of Glaser et al. (1981) that the "MDL is considered operationally meaningful only when the method is truly in detection mode, i.e., [the] analyte must be present." is problematic. For methods under which a signal is generated from blanks, this is not at all necessary, For cases in which the blank does not generate a signal due to instrumental limitations (such as inability to find the peak to integrate), one must generate the MDL using positive concentrations, and the MDL should not increase the CRV much over what would be obtained using true blanks if that were possible. Otherwise, blank samples are fine.

Response: EPA agrees and has addressed this issue by revising the MDL procedure to allow use of blanks in which the analyte is present for determination of an MDL.

Regulatory levels (DR-10)

Comment: Obviously, levels of a toxic substance cannot easily be regulated below the level at which there is an instrumental response (i.e., a signal is generated). All environmental measurements should be reported as measured, and should only be reported as non-detects if there is no instrument response. If a value is generated by the instrument, it should be reported, with an indication of what the estimated standard deviation is, and whether the measurement shows the concentration to be non-zero (that is, whether the signal is above the CRV).

Response: In principle, EPA agrees with the reviewer, and is aware that several organizations, including the European Union and others, are developing procedures for estimating the uncertainty associated with measured results. If successful, such an approach would eliminate many of the data censoring concerns discussed in Section 3.3.2 of the Assessment Document. Given the difficulty in achieving consensus on an appropriate means of establishing a detection limit, however, EPA believes that it would be extremely difficult to obtain consensus on an appropriate means for estimating the uncertainty associated with each result measured on each environmental sample. In addition, analytical chemists have used and believe that they understand a quantitation limit to mean the lowest concentration at which an analyte can be identified and determined with some degree of certainty, and most laboratories are reluctant to report results with the higher error at a detection limit, as compared to a quantitation limit. Therefore, EPA prefers to monitor developments by the

EU and others on this subject, and if appropriate, re-evaluate this issue if and when it becomes widely accepted by the environmental laboratory communities. In the meantime, EPA believes that the traditional approach of defining a quantitation limit at some level above the detection limit provides a data user with a reasonable degree of confidence in the reliability of the measured value without requiring that individual estimates of uncertainty be developed and reported. Therefore, EPA has not revised Criterion 5 to reflect the changes suggested by this peer reviewer.

MDL is appropriate CET (DR-11)

Comment: For substances for which the toxic level is well below the CRV, the compliance evaluation threshold should be at the critical value (in one of its implementations such as the revised MDL).

Response: In principle, EPA agrees with the reviewer. However, EPA has suggested that the compliance evaluation threshold (CET) be set at the ML because of added reliability of measurements at the ML. States that wish to be more protective of the environment may set the CET to the MDL or other limit are allowed to do so if they believe is will be more protective.

Prediction and tolerance intervals (DR-13)

Comment: Tolerance intervals are inappropriate for environmental monitoring. The main issues here are 1) is the true concentration greater than some specified safe of action level, with sufficient confidence, and 2) what interval of possible concentrations is consistent with one or a series of measurements, with a specified degree of confidence. Both are statements about a given sample or series of samples, and not about the hypothetical variability of future estimates. Suppose that one has a sample of 10 observations with mean concentration of 1 ppb and standard deviation of 0.5 ppb. The estimated 99% CRV is (2.326)(0.5) = 1.2 ppb. One may choose to use a *t*-score instead of a normal score so that the chance that a future observation will exceed this level is in fact 99%. In this case, the CRV estimate would be (3.250)(0.5) = 1.6 ppb. This does actually correspond to a prediction interval for future observations from a zero concentration sample.

If one asked instead for a 95% confidence interval for the .99 percentage point of the true distribution of measurements (assuming normality) when the true quantity is zero, this can be calculated approximately using a chi-squared distribution and covers the interval (0.9 ppb, 2.4 ppb). It does not, however, make sense to use 2.4 ppb as a threshold, since the chance of a future observation exceeding 2.4 ppb when the true mean concentration is 0 is about .0005, far smaller than the intended false-positive rate of .01.

Response: EPA agrees.

Spike to MDL ratio (DR-17)

Comment: Use of as much as five times the CRV for the spike concentration could be problematic. Inflation of the MDL by using a spike at the CRV is only 25% for a method with a coefficient of variation of 20% (this and other calculations here are done with the Rocke and Lorenzato 1995 variance function assuming a sample size of 7). A spike concentration of 3 times the CRV inflates the MDL to a value 140% higher, which even there may be tolerable. Use of a value 5 times the CRV gives an inflation of over 280%. Thus if the true CRV is 1 ppb, then the use of 1, 3, and 5 times the CRV for spike concentrations in determining the MDL gives likely values of 1.2 ppb, 2.4 ppb, and 3.8 ppb, respectively. These number were determined as follows: Let $V(y) = a^2+b^2\mu^2$. Then the expected MDL if blanks were used is approximately ta, where t is the appropriate t-statistic. If spikes at kta are used, then the variance at that level of μ is $a^2+(ktab)^2$, and the approximate estimated MDL will be t times the square root of this quantity, so that the ratio of the MDL with blanks to the MDL at spike level $\mu = kta$ is $\sqrt{[1+(ktb)^2]}$. Thus, I would recommend that the procedure be altered to use concentrations that are no more than 3 times the detection limit, and perhaps to permit concentrations lower then the CRV, including possibly blanks.

Response: EPA agrees with the reviewer's concern, but has found that, in practice, the disadvantages associated with limiting the spike-to-MDL ratio to three outweigh the benefits. Specifically, EPA initially required that all laboratories participating in the EPA's Multi-technique Variability Study (the Episode 6000 Study) achieve a spike-to-MDL ratio of three. After attempting to meet this requirement, two of laboratories reported that a large number of iterations would be required in order to achieved a the lower spike-to-MDL ratio. Given the costs associated with each iteration, EPA returned to a ratio of 5 as was stated in the original MDL procedure. EPA is retaining a ratio of 5 in the revised procedure being proposed.

Responses to Comments Submitted by Walter Piegorsch

Peer-review was complete and sound (WP-1, WP2-7, WP2-13)

Comment: The peer review version of the Assessment Document is well-organized and intelligently thought-out. It has strong scientific merit and establishes a good baseline from which further discussion and debate may continue on the important issue of detection limits and quantification of contaminants in the nation's water supply. The evaluation criteria in Chapter 4 seem reasonable. The description of the IUPAC/ISO detection limit and the introduction to quantitative assessment of the ML were well presented. EPA's willingness to consider other detection and quantification approaches is admirable.

Response: EPA thanks the reviewer for the support.

MDL is not an interval estimator (WP-3, WP2-2)

Comment: Accepting the Assessment Document's interpretation of the MDL as a "general purpose version of Currie's critical value (L_C)," I am concerned that the operational definition MDL = $t_{0.99}(\mathrm{df})S$, where df is an appropriately-chosen value for the degrees of freedom and S is an associated root mean square, does not correspond to an appropriate form of interval estimator.

Response: The MDL is not intended as an interval estimator but rather a value analogous to a Currie critical value (CRV; L_C). EPA has removed all references to interval from the revised version of the MDL procedure being proposed as a result of this assessment.

Correct terminology (WP2-3)

Comment: The peer review version of the Assessment Document, and I, should be more careful in the use of statistical terminology. We both refer often to confidence "intervals," when in fact the quantity of interest is a confidence limit — or tolerance limit, etc. — on some underlying parametric quantity.

Response: EPA will try to be consistent with reference to limit, and has revised the Assessment Document accordingly.

Critical value not detection limit (WP2-6)

Comment: There seems to be a fair amount of confusion on the issue in the analytical chemistry literature. The bottom line from my reading of the Assessment Document is that, in effect, we are calculating an L_C, but using terminology that makes some readers think it's an L_D . The Agency should put forth an effort to overcome this confusion in terminology.

Response: EPA agrees that the terminology is inconsistent, and to help overcome confusion arising from the terminology has added the following sentence to the MDL definition, "*The MDL is calculated from replicate analyses of a matrix containing the analyte and is functionally analogous to the "critical value" described by Currie (1968, 1995) and the Limit of Detection (LOD) described by the American Chemical Society (Keith et al.1980, MacDougall et al. 1983).*"

Error in MDL equation in the draft Assessment Document (WP-4, WP2-9)

Comment: The definition of S^2 given on p.5-2 of the peer review version of the Assessment Document is in error. I assume that this a typographical error and not a more serious misinterpretation of statistical principles.

Response: EPA thanks the reviewer for pointing out this typographical error and has corrected it in the revised version of the Assessment Document included in the Docket for today's proposal.

MDL confidence interval is fallacious (WP-5, WP2-10, WP2-14)

Comment: The idea of building a confidence interval for the MDL is fallacious because a 95% confidence interval for the MDL cannot be calculated. The suggestion that MDL represents a 95% confidence interval is spurious.

Response: Although EPA asserts that a confidence interval can be calculated for an MDL, we have deleted all references to a confidence or other interval from the MDL procedure because, operationally, these intervals serve no purpose.

Prediction and Tolerance Intervals (WP-6, WP2-4)

Comment: Some authors even argue that the Glaser *et al*. model and definition of the MDL does not even produce a valid confidence, prediction, or tolerance interval. Rather than join the fray here, however, I suggest the following reconsideration: Accepting the Assessment Document's argument on p.3-25 that the practical value of tolerance limits for this sort of analyte detection is limited, one naturally thinks to view the MDL as a prediction limit. But, as Gibbons (1994, p.98) points out, a single-use prediction limit of such a form should contain an additional term, viz.

$$t_{0.99}(\mathrm{df})S.$$

I emphatically encourage the EPA to revisit its definition of MDL with this consideration in mind.

Response: EPA appreciates and has considered the suggestion. EPA wishes to note that, although there have been suggestions in the literature concerning prediction and tolerance intervals, neither the Currie critical value (CRV), the ACS limit of detection (LOD), nor the EPA MDL have historically included allowance for a prediction or tolerance interval. One use of the

MDL (and other detection limit approaches that are based on the standard deviation of replicate measurements) is to characterize the performance of a method or laboratory. Because there is no intent to test a future MDL or other measurement against the value estimated or determined, a prediction or tolerance interval is inappropriate. EPA acknowledges that a prediction or tolerance interval may be applicable if the MDL were to be applied on a routine basis; e.g., for regulatory compliance. EPA refers the reader to comment DR-13 above, provided by another peer reviewer, for an excellent explanation of why prediction or tolerance intervals are inappropriate for environmental monitoring programs.

False negatives (WP-7, WP2-5)

Comment: The single most problematic issue in developing a detection limit is that of correction for false negatives. For the generic problem of detecting a chemical analyte, the incorporation of false negatives should be afforded greater importance than I think the peer-review Assessment Document provides.

Response: EPA agrees that the issue is problematic and attempted to address the issue in the Assessment Document.

MDL not adjusted for outliers (WP2-11)

Comment: It is good to mention that the MDL procedure is not adjusted for outliers, since this sort of subtlety could escape the casual reader.

Response: EPA appreciates the positive feedback. EPA also notes that, based on feedback from other reviewers of the draft Assessment Document, EPA is proposing revisions to the MDL procedure that will allow removal of a maximum of one outlier in the revised MDL procedure and include guidance on the outlier removal processes. EPA has revised the Assessment Document to further discuss this issue.

Calibration design (WP-9)

Comment: There is also the rather strong argument that instead of L_C calculation from a single concentration design, use of calibration designs is considered to be more efficient in terms of deriving effective detection limits from the data.

Response: EPA is not opposed to the “calibration” design (also referred to as the “concentration” design) on scientific principles; EPA is opposed to use of the calibration design to establish detection and quantitation limits on practical and economic grounds. However, EPA is proposing to allow use of detection and quantitation limits developed from alternate approaches and procedures, including the "calibration" or "concentration" design.

IDE procedure is complex (WP2-12)

Comment: It is worth emphasizing that the IDE procedure as outlined is so complex as to make simple determination of error rates associated with it untenable.

Response: EPA has incorporated this reviewer's comment into the revised version of the Assessment Document.

Draft Assessment Document favors MDL (WP-10)

Comment: The evaluation criteria in the peer-review version of the Assessment Document seem reasonable at first reading. I do not think any of them should be eliminated, and I do not have any concrete suggestions for addition. In passing, however, I should note that as I continued through the chapter, I found it perchance-less-than-coincidental that the (revised) MDL and ML approaches seemed to satisfy the criteria so readily, and that most of the other approaches were found wanting. (A cynical reader might view this as a contrivance that elevates the MDL and ML at the expense of the other methods, and perhaps the EPA may wish to proceed with caution in this area.)

Response: EPA acknowledges that a reader of the draft Assessment Document would conclude that the MDL better satisfies the evaluation criteria than the other detection and quantitation limit approaches evaluated. The reason is that the evaluation criteria were constructed around EPA's needs and uses for detection and quantitation limits and approaches, and these needs have scientific, practical, cost, and regulatory components. These needs have not changed substantially since the 1980s when MDL procedure was developed and adopted to support them. EPA believes that if other approaches had been developed to suit Agency needs, these approaches would have better satisfied EPA's evaluation criteria. As an example, if the ISO/IUPAC critical value had a well-defined procedure for its determination, it is likely that the critical value would have met the evaluation criteria equally well to the MDL.

Additional data sets (WP-13, WP2-15)

Comment: I would encourage the EPA to expand its search and consider as many additional data sets as can be acquired in a reasonable period of time. I do not have at my disposal any new data sets, nor am I working with anyone currently who does. I cannot give EPA any new sources of data.

Response: EPA continues its search for definitive data sets that may further assist in resolving the detection/quantitation limit issue and is soliciting additional data sets in the proposal.

New interlaboratory study (WP-14, WP2-16)

Comment: The issues of intralaboratory and interlaboratory variation are quite important, and I applaud the Assessment Document for its consideration of them. While reasonably addressed, I would encourage that EPA undertake, commission, or actively abet a formal interlaboratory study, building on the success of the Method 1638 Interlaboratory Validation Study. The recognition that multiple components of variation can exists in calculating L_C (or any other form of detection/decision limit), is an important one, and such calculations must be based on appropriate variance components for the model under study. A large, carefully-conducted interlaboratory study would make a major contribution towards understanding and quantifying these components for use in future detection limit calculations.

Response: When EPA discussed the objectives for the Multi-technique Variability Study (the Episode 6000 study) in 1996, the purpose was to create a database that would help answer questions concerning detection and quantitation limits. The study was conducted using 7 replicates at 16 concentration levels and 11 of the most commonly used analytical technologies. EPA considered conducting the study in multiple laboratories at the time but the additional factor of 6 to 10 in cost made this approach cost-prohibitive. EPA believes, however, that the Episode 6000 study shows that detection limits can be variable and that some judgement will always be required in evaluating detection limit determinations; and that there is no procedure that is applicable to all analytes in all methods under all circumstances. If an organization is willing to conduct more extensive studies between proposal and promulgation of a final rule, EPA would appreciate additional data so long as it doesn't compromise the deadline for the final rule.

Composite sampling (WP-15, WP2-1)

Comment: Aside from the comments given above, I have no further improvements to suggest. I do have one general question, however: has EPA studied the use of composite sampling methodology (primarily from a statistical perspective) for application to MDL or ML determination?

Response: EPA has considered composite sampling in its various regulatory programs and other data gathering, and has published procedures for composite sampling in regulations and analytical methods. Analysis of composite samples with respect to detection and quantitation is not fundamentally different from analysis of individual samples. In fact, no advantage is gained with respect to detection and quantitation by compositing samples. Compositing a number of samples with concentrations below detection levels with one or a few samples above detection levels can cause the composite sample to have a concentration below detection. Increased

replicate analysis of the same sample would be a better means of improving a detection limit estimate.

A new approach should be assessed in same way as those presented (WP2-8)

Comment: If a revised MDL or some other new limit calculation is adopted, it should be assessed in the same way that the various approaches have been assessed.

Response: EPA is not proposing a new approach in today's rule, but is proposing modifications to the existing MDL and ML approaches. These modifications are included in the assessment detailed in the Assessment Document that supports this proposal. EPA also is soliciting comments on a new approach recently suggested by the Inter-Industry Analytical Group (IIAG). EPA agrees that if comments suggest further consideration of the IIAG approach, it (and any other new approach) must be assessed in the same way that the various approaches have been assessed in the Assessment Document.

Responses to Comments Submitted by Dallas Wait

Data Quality Act (PL 106-554), (DW-1)

Comment: The Data Quality Act mandates that the Office of Management and Budget (OMB) issue guidance to federal agencies for "ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by federal agencies." The construct of EPA's draft Assessment Document is consistent with the spirit of this act.

Response: EPA thanks the peer reviewer for the recognition.

Daubert (DW-2)

Comment: EPA should be applauded for invoking the Daubert factors (testing and validation, peer review, rate of error, and general acceptance in the scientific community). Use of the Daubert approach is defensible and should give the resultant consensus document long-term standing.

Response: EPA appreciates the positive feedback and agrees that the Court's definition of scientific validity has been helpful in evaluating alternatives.

Resolution of detection/quantitation issues (DW-3)

Comment: Overall, I believe the Assessment Document effort is a rigorous, open and honest attempt by EPA to resolve a technically and operationally difficult matter in a manner fair to all sides.

Response: EPA appreciates the peer reviewer's opinion.

Literature search (DW-4)

Comment: The thoroughness of the on-line search for relevant documents provided in the Reference Section and Appendix A of the Assessment Document was impressive. However, on-line searches often don't capture information contained in text and reference books. This area of information may not have been adequately addressed. EPA's literature search was extensive, regardless of my suggestions to examine some other sources.

Response: EPA has included references to text and reference books in the revised version of the Assessment Document in response to this suggestion.

Perception of detection/quantitation approaches by other Federal and State agencies (DW-5)

Comment: EPA appears to have closely examined the detection and quantitation approaches by professional organizations. Has EPA rigorously examined how these approaches are perceived and implemented by other Federal and State agencies (e.g., USGS, NRC, FDA, DOD, DOE)? For example, Chapter 19 of the recently published draft document entitled "Multi-Agency Radiological Laboratory Analytical Protocols Manual" (MARLAP) discusses detection and quantitation issues. It would be useful for EPA to tabulate the approaches used by federal and state agencies in the Assessment Document.

Response: In response to this comment, EPA has added a new section on "Approaches Advocated by Other Governments" to Chapter 2 of the Assessment Document (see Section 2.3.4). EPA did not tabulate use by federal and state agencies because such a tabulation would be a formidable task that is unlikely to yield new information or improve the assessment. For example, EPA found more than 140 uses by more than 50 organizations in developing the Environmental Monitoring Methods Index (EMMI) in 1995. In many instances, different organizations use the same term to describe different things. In other instances, different organizations use different terms to describe the same thing. Also, EPA has found that nearly all of the state and federal "uses" are merely different applications (i.e., reporting thresholds, permit limits, certification criteria, etc.) of the basic approaches that EPA evaluated or are only slightly different variants of these approaches. For example, the U.S. Geological Survey has adopted something called a long-term MDL that is based on many of the same fundamental assumptions as the MDL. Therefore, rather than tabulating uses by different organizations, EPA focused the assessment on the most widely used and recognized detection and quantitation limit approaches, and incorporated a consideration of state and local implementation of these approaches into it's evaluation criteria (see Evaluation Criterion 6)

Case law (DW-6)

Comment: Another source of information may be Case Law. Has EPA examined whether there is a legal record detailing EPA (and others, e.g., NEIC, DOJ) opinions on these matters? If so, their opinions and those of the Court should be acknowledged.

Response: EPA has not done a rigorous search of case law for litigation on detection/ quantitation limit approaches in general, and has instead focused its assessment on the scientific basis for many approaches and practical aspects of implementing those approaches to meet EPA needs under the Clean Water Act.

Justification for seven replicates in the MDL procedure (DW-7)

Comment: EPA should justify why seven replicates were chosen to determine MDLs rather than six, eight, or some other number.

Response: During development of the MDL procedure, the use of seven replicates was selected in consultation among several EPA analysts and statisticians based on (1) the number that would be not overly burdensome to measure in the laboratory and (2) the point at which the multiplier is approaching an asymptote in Student's *t*-test. There is no absolute justification for exactly seven replicates, in the same way that there would be no justification for exactly six or exactly eight replicates. In response to this peer reviewer's suggestion, EPA has modified Section 2.2.1 of the Assessment Document to include this explanation.

Matrix effects (DW-8)

Comment: Matrix effects are an extremely critical element to be considered when generating MDLs. As EPA notes, since each environmental sample is unique, it would be impossible to conduct a MDL study on each. The best means of dealing with this reality is by employing on a project by project basis a graded approach to verifying MDLs. The EPA data quality objectives (DQO) process is an efficient mechanism for addressing the variability of MDLs between different matrices.

Response: EPA agrees that the DQO process would be an effective means for dealing with the variability of MDLs between different matrices.

Reference matrices (DW-9)

Comment: EPA states that it believes that reference matrices should be used to establish method detection and quantitation limits. Has EPA considered establishing a repository of "typical" matrices where low background concentrations of contaminants are thoroughly characterized similar to the National Institute of Standards (NIST) Standard Reference Materials (SRMs)? If laboratories had the option of evaluating MDLs using matrices similar to samples they were studying (e.g., POTW wastewater, salt water, river sediment, pond sediment, clay), this would give labs an option in demonstrating their analytical capabilities in a fashion comparable to other labs. Use of these low level matrices would be determined during the DQO process.

Response: EPA has considered, at various times, providing reference matrices for performance evaluation and other purposes. Difficulties include (1) the stability of aqueous samples and the holding times necessary to assure stability, (2) the argument that no reference matrix would represent all possible matrices, even within a given industrial category, (3) making and

preserving a reference matrix requires extensive study, as evidenced by NIST's experience, (4) the costs involved in developing and maintaining a repository of reference matrices, and (5) the potential conflict with NIST and with non-governmental organizations that provide reference matrices. In addition, the ultimate use of the reference material would be for performance evaluation, a situation likely to bring contention over the reliability of the reference material. Given all of these difficulties, EPA believes that it is better to have an organization with experience in developing SRMs, such as NIST provide reference materials. In response to this reviewer's comment, EPA has added a section titled "Repository of Reference Matrices" to the Assessment Document supporting today's proposed rule (see Chapter 3, Section 3.1.3.2).

Allowance for use of a different detection/quantitation approach by a laboratory (DW-10)

Comment: Section 3.2.4 of the Assessment Document discusses, in part, the option of using performance standards over prescriptive standards, which would allow laboratories and others the freedom to use a variety of different approaches to establish limits. Although theoretically this sounds agreeable, operationally this would be a nightmare and comparability, a QA tenet, would be jeopardized. I'm not in favor of this approach.

Response: EPA is committed to flexibility in analytical measurements, and EPA's water programs have favored a reference method or procedure with QC acceptance criteria for establishing a performance benchmark. However, EPA does not want to force other organizations to use the EPA approach for detection and quantitation because other approaches may also be suitable for CWA applications. Therefore, in this instance, EPA believes that it is better to standardize on one pair of procedures (detection and quantitation) and continue to accept methods that use alternate procedures when the method provides the sensitivity and performance needed for use in CWA programs. For example, EPA would continue to accept a Standard Method or ASTM method that specifies a different type of detection or quantitation limit, provided it meets EPA's regulatory needs.

Sources of variability (uncertainty) (DW-11)

Comment: Overall, the analytical chemistry, CWA regulatory issues, and statistical issues presented in Section 3 of the draft Assessment Document are comprehensive. In Section 3.3.1, the discussion on sources of variability could be enhanced to address the impact of variability at the MDL and how this variability impacts data use.

Response: EPA agrees and, in response to this suggestion, has enhanced the discussion in Section 3.3.1 of the Assessment Document.

Measurement quality over the life of a method (DW-12)

Comment: Although informative, Section 3.1.4 of the Assessment Document, which discusses measurement quality over the life of a method, could probably be deleted without hurting the integrity of the Chapter.

Response: After carefully considering this comment, as well as comments from other peer reviewers, EPA has decided to retain the section (which now appears as Section 3.1.5 of the revised Assessment Document). EPA discussed the issue of measurement quality over the life of an analytical method in the context of detection and quantitation because method performance, including the ability to detect and quantify to lower levels, improves with time. EPA believes this is an important point that is worth making because detection and quantitation limits published in the methods may not reflect current method and laboratory capabilities.

Adequacy of evaluation criteria (DW-13)

Comment: All of the criteria used by EPA are pertinent to the evaluation of viable detection and quantitation limit methods. The explanations for each criterion are reasoned and persuasive. I would not remove any criteria from the evaluation process. No other evaluation criteria are apparent.

Response: This comment was consistent with comments made by other peer reviewers. EPA appreciates the positive feedback.

Validity of assessment (DW-14)

Comment: The thorough evaluation process used by EPA is excellent! A comprehensive and open discussion was performed for all five detection limit approaches and four quantitation limit approaches. These discussions fairly debate the pros and cons of each approach.

Response: EPA thanks the peer reviewer for the comment.

Clarification of Step 1 of the MDL procedure (DW-15)

Comment: In Section 5.1.1.2.1 of the Assessment Document, EPA astutely notes that many people complain that MDLs can vary depending on spike levels used, based on the mistaken assumption that spike levels may be arbitrarily selected. I have witnessed this same complaint numerous times. EPA also properly notes that Step 1 of the MDL procedure specifies a number of criteria that must be met in selecting spike levels. Apparently many chemists just don't get it. It would be advantageous for EPA to embellish Step 1, possibly with examples, to make the requirement clearer.

Response: EPA agrees and is proposing revisions to the MDL procedure. Among these clarifications are more specific instructions at Step 1.

Alternative detection/quantitation procedures (DW-16)

Comment: The Assessment Document is not clear in what options EPA is considering and what alternative approaches and procedures EPA will accept. Any flexibility in approaches and procedures could lead to further litigation.

Response: EPA has included the following statement in the MDL procedure proposed today: "An alternative procedure may be used (e.g., from a voluntary consensus standards body) to establish the sensitivity of an analytical method, provided the resulting detection limit meets the sensitivity needs for the specific application." This statement allows EPA to continue to accept methods from VCSBs with detection limits other than MDLs.

EPA does not believe that flexibility in detection and quantitation approaches and procedures would necessarily result in further litigation. EPA has already approved hundreds of methods developed by other organizations, and many of those methods cite sensitivity ranges that are not based on the MDL. In considering methods for approval under the CWA, EPA considers the regulatory needs associated with the method and all available performance information, including information concerning method sensitivity. EPA is unlikely to approve a method if it will not support regulatory needs. If supporting information is not adequate to justify proposing use of the method, EPA requires additional information from the method developer prior to approval.

Better correlation between Table 6-1 and the text (DW-17)

Comment: A better correlation between the findings in Table 6-1 of the Assessment Document and the associated text would be useful. Within Table 6-1, it would also be useful to reference where in the Assessment Document many of the statistics were derived. Also, the revised MDL procedures presented in Appendix D should be mentioned.

Response: EPA agrees and has made the recommended changes to the revised version of the Assessment Document released today.

Additional data sets (DW-17)

Comment: EPA has requested additional data sets. During the 1980s numerous interlaboratory method evaluation studies were conducted by EPA's ORD group in Cincinnati, some of which may have looked at detection limits. Has EPA examined any of their historical work for pertinent MDL information? Also, as I recall, George Stanko of Shell presented a fairly large study challenging EPA's detection limits for volatile organics in water

at EPA's annual Analytical Symposium in Norfolk, Virginia? Has EPA petitioned large trade associations, such as the American Petroleum Institute (API), about detection and quantitation studies they may have sponsored? Personally, I am not aware of any additional detection and quantitation limit data sets that may be of value to EPA.

Response: The interlaboratory method validation studies conducted by EPA's laboratory in Cincinnati were not directed at determining detection and quantitation limits but rather at characterizing precision and recovery across the analytical range. Regarding studies by the late George Stanko of Shell, EPA reviewed the minutes of the Annual Analytical Symposia and found the papers by George Stanko, but the papers did not include data that could be used to evaluate the various detection and quantitation approaches. Regarding data sets from studies sponsored by trade associations, EPA has requested such information in discussions with the Electric Power Research Institute (EPRI), the Inter-industry Analytical Group, and others. The Petitioners and Intervenor provided EPA with a list of databases for consideration. This list, and EPA's decision regarding each of the recommended databases, is discussed in the revised Assessment Document. In addition, EPA is soliciting any appropriate data sets in the proposal.

Inter- vs intra-laboratory issues (DW-18)

Comment: Use of inter-laboratory measurements is important for a general understanding of the laboratory communities' capabilities, but is not as relatable to the issues that EPA must consider in support of a permittee's CWA requirements. Intra-laboratory measurements are more practical. EPA's approach between inter- and intra- studies is balanced and reasonable.

Response: EPA appreciates the support.

Improvements to detection/quantitation procedures (DW-19)

Comment: The MDL and ML approaches evaluated in the Assessment Document are shown to be technically sound and practical. The revised MDL procedure provided in Appendix D is streamlined and more intelligible than the previous version, although a reexamination of Step 1 to aid chemists in the spiking level requirement may be warranted. The detection and quantitation approaches I'm aware of have already been adequately "fleshed" out by EPA.

Response: EPA appreciates the positive feedback and has added detail to the spiking procedure, as the peer reviewer suggests.

Recommendations for improvements to the Assessment Document (DW-20)

Comment: Although the Assessment Document is necessarily long and dense with information, it is well written and flows logically. I would not make any structural changes to the document.

Response: EPA appreciates the feedback, and has retained the overall structure in the revised version of the Assessment Document.

EPA's Quality System (DW-21)

Comment: Since the Assessment Document addresses fundamental quality assurance issues, I'm surprised that there is no acknowledgment or reference to EPA's Quality System. EPA may want to reexamine the Assessment Document and update as appropriate to remain consistent with Agency directives.

Response: In response to this comment, EPA Office of Water staff searched the Quality System documents for specific information regarding detection and quantitation, and could not find a direct reference. EPA believes, however, that the assessment is consistent with policies embodied in EPA's Quality System.

Acronyms and abbreviations (DW-22)

Comment: A listing of acronyms and abbreviations would be useful.

Response: EPA agrees and will add a list to the final Assessment Document.

Typographical errors (DW-23)

Comment: A list of typographical and grammatical errors is included.

Response: EPA thanks the commenter for the list and has corrected the typographical and grammatical errors.

Information that EPA could add to the Assessment Document (DW-24)

Comment: A list of additional references is included.

Response: EPA thanks the commenter for the list and has added the references.

Responses to Comments Submitted by Marcus Cooke

EPA should consider European approaches to detection and quantitation and treatment of analytical data (MC-1, MC-5)

Comment: EPA could consider method verification, data reliability, and detection approaches that have been developed by the European Union (EU). One example is the application of the operational equivalent of EPA Method 1613B for chlorinated dibenzo-p-dioxins and dibenzofurans (CDDs/CDFs). Vegetable foodstuffs must be tested at levels more than an order of magnitude below the method detection limit (MDL) of 1-5 parts-per-trillion (ppt) in EPA Method 1613B. EU regulators applied an "Upper Bound" reporting limit where non-detects are found, using the EPA MDL for each analyte. This forces laboratories to achieve levels available with modern instrumentation, otherwise the "Upper Bound" reporting level is above the regulatory compliance level and the data (or foodstuffs) are rejected.

EPA may want to consider recent advances in the statistical treatment of analytical method data that has evolved in Europe, for 3 reasons: (1) The EU is conducting the largest trace chemical analytical program in the world (EPA is the record holder with the Contract Laboratory Program, $1.5B+); (2) The EU has applied an operational equivalent of EPA Method 1613B to a regulatory program that screens all food and animal feed used, produced, or imported; and (3) The subject EU program has developed practical solutions to applying modern ultra-trace measurements (and statistical verification) in a legally-based, widely-applied testing program.

Response: EPA has revised Chapters 2, 3 and 4 of the Assessment Document to address these suggestions. It is important to recognize that the Upper Bound approach described by the reviewer is directed at reporting limits, rather than detection or quantitation limits, and that in adopting this approach, the EU has essentially 1) accepted the MDL concept, 2) used the MDL as a reporting limit, and 3) utilized the relatively low cost and ease of the MDL procedure as a tool for encouraging improvements in measurement technology. EPA agrees that this approach, which yields a "worst-case" (i.e., highest possible) estimate of the pollutant concentration, can serve as a useful tool for encouraging the analytical and regulated community to pursue measurements at the lowest levels necessary to protect human and ecological health. EPA believes such an approach may be useful in the context of establishing permit limits, but that the approach does not shed new light on the issue of establishing detection and quantitation limits.

A reference material would aid in evaluating "trueness" (MC-2)

Comment: Many EU procedures have a "trueness" criterion. This is accuracy determined by percent recovery of an accepted reference material. In order to incorporate trueness into an EPA method validation study, an appropriate

reference material would need to be developed ahead of time and included in the study.

Response: For method validation and periodic testing, EPA uses reagent water as the reference material because reagent water is available to all laboratories. Where possible, EPA requires use of a reference material for periodic method performance verification. For example, with each analytical batch of 20 or fewer samples, Section 9.5 of EPA Method 1631E requires analysis of mercury spiked into reagent water (the "ongoing precision and recovery" (OPR) sample), and Section 9.5 requires analysis of a quality control sample (QCS). The QCS can be a Standard Reference Material (SRM) from the National Institute of Standards and Technology (NIST) or other material obtained from a different source than the calibration standard and OPR standard. Recovery of the OPR must be within the QC acceptance criteria in EPA Method 1631 and recovery of the SRM or other QCS must be within the limits specified by the organization responsible for providing the material.

While EPA agrees that the use of such reference materials can be of value in a validation study, EPA notes that the universe of SRMs or certified reference materials (CRMs) for matrices and analytes relevant to the Clean Water Act is limited. More importantly, few, if any, SRMs are available at concentrations in the region of interest for establishing detection and quantitation limits of a given method. EPA has added a discussion of these issues to Chapter 3 of the revised Assessment Document.

European protocols give tools for uncertainty evaluation (MC-3)

Comment: The Eurachem Guide *Quantifying Uncertainty in Analytical Measurement*, Second Edition (QUAM:2000.P1), provides guidelines to evaluate uncertainty in analytical measurements. These elements were detailed in a presentation at the 22nd International Symposium on Halogenated Environmental Organic Pollutants and POPs [persistent organic pollutants] (the "22nd International Symposium). United Kingdom Valid Analytical Measurement Programme (VAM) Project 3.2.1 *Development and Harmonisation of Measurement Uncertainty Principles* (LGC/VAM/1998/088, January 2000) protocols give additional tools for uncertainty evaluation.

Response: In response to this suggestion, EPA reviewed the Eurochem and VAM documents and modified Chapters 2 and 4 of the Assessment Document to describe EPA's findings. The European Union (EU) guidance advocates reporting all results along with an estimate of the uncertainty associated with each value. In its discussion of the issue, the EU indicates that use of the term "limit of detection" only implies a level at which detection becomes problematic and is not associated with any specific definition. Instead, the EU focuses its attention on ways to estimate uncertainty, basing

its approach on the ISO Guide to the Expression of Uncertainty in Measurements (1993). The EU also notes, however, that the use of uncertainty estimates in compliance statements and the expression and use of uncertainty at low levels may require additional guidance. The VAM document, published by the United Kingdom (UK), uses a similar approach. Because both of these approaches focus on estimating uncertainty rather than at establishing or defining limits for detection and quantitation, EPA does not believe they warrant extensive evaluation as part of the Agency's assessment of approaches for establishing detection and quantitation limits. Moreover, given the difficulty in achieving consensus on an appropriate means of establishing a detection limit, EPA believes it would be extremely difficult to obtain consensus on an appropriate means for estimating the uncertainty associated with each result measured on each environmental sample. EPA will continue to monitor developments by the EU and others on this subject, and if appropriate, re-evaluate this issue if and when it becomes widely accepted by the environmental laboratory community.

Contributions to uncertainty (MC-4)

Comment: In another presentation at the 22nd International Symposium, contributions to measurement uncertainty for CDD/CDF analysis of food and feed were presented. The major contributions were from precision, trueness (bias), purity (not defined, but presumably the purity of the reference material used in the analysis), and [sample] homogeneity.

Response: EPA is familiar with the presentations and agrees with identified contributions to measurement uncertainty. Sources of uncertainty are useful for understanding the analytical process so that analytical methods can be refined in an attempt to reduce uncertainty and thereby lower detection and quantitation limits.

Quality control (MC-6)

Comment: Chapter 3 of the Draft Assessment Document addresses 20 technical elements, and Chapter 6 addresses 6 directed issues. EPA may want to consider the additional issues of quality control (QC) and use of reference materials. These issues may have a significant effect on the reliability of data produced at ultra-trace levels, whether to determine the initial presence of an analyte like mercury, or reliably apply regulations at a discharge limit.

Single laboratories, working independently, start from scratch each time they perform a method. QC should be sufficient to insure reliability in single, isolated determinations of small sample sets, as well as in large commercial laboratories performing many tests.

Response: The methods approved for use under the Clean Water Act include a standardized suite of quality control procedures and analyses that are

designed to ensure that the analytical results can be assessed and their reliability can be determined, regardless of the size of the laboratory performing the analysis. Where practical, analyses of reference materials form a portion of those QC procedures. However, as noted earlier, the universe of relevant SRMs in limited, and their relevance to establishing detection and quantitation limits also is limited.

However, in response to this suggestion, EPA has revised Chapter 3 of the Assessment Document to include additional discussion of QC and reference materials in Chapter 3 (see Section 3.1.3.2, Repository of Reference Matrices and Section 3.1.5, Measurement Quality of the Life of a Method).

Use of a reference material for calibration (MC-7)

Comment: Calibration required in Method 1631B could be enhanced by use of a reference material which contains a "real world" matrix, and also mercury forms known to exist in natural samples.

Response: Enhancements to EPA Method 1631 are outside the scope of the assessment that is the subject of this peer review. However, EPA will consider the suggestion during any future revision of Method 1631. For the record, Section 7.7 of EPA Method 1631E (and previous revisions) requires that the stock mercury standard be the NIST-certified 10,000-ppm aqueous Hg solution (NIST-3133).

Mercury forms, species, and compounds (MC-8)

Comment: Elemental mercury in nature often converts to Cinnabar or meta-Cinnabar, forms of mercuric sulfide. These are very stable, innocuous forms of mercury. Ambient samples can also contain organomercurials that have elevated human toxicity. EPA might consider a demonstration study to show how "safe" or "unsafe" mercury forms are oxidized by BrCl, and are subsequently measured by Method 1631, especially at low concentrations near the limit of detection.

Response: EPA appreciates the suggestion. However, further evaluation of EPA Method 1631 is beyond the scope of the assessment that is the subject of this peer review.

Cost and ease-of-use (MC-9)

Comment: The complex theoretical treatments defined in the Assessment Document and the resultant additional analyses required in regulatory applications of Method 1631B may produce significant cost due to new supporting analyses needed to demonstrate detection and data reliability. The Assessment Document does not adequately address cost to users. Considerations of ease-of-use and cost should be included in any final revisions arising from

this process. Cost issues should be fully addressed in summary reviews of the Assessment Document.

Response: EPA agrees that the "complex theoretical treatments" involved in some of the approaches to establishing detection and quantitation limits would raise the costs of establishing such limits. Such costs would be incurred by the developer of the method (e.g., EPA or another organization). EPA is even more concerned about increasing the costs incurred by individual laboratories and, therefore, passed on to their clients (data users) if each laboratory was required to duplicate the "treatments" involved. Therefore, in its assessment, EPA favored approaches that could be readily applied in a single laboratory, either to establish or verify the detection and quantitation limits associated with an analyte.

In response to this comment, EPA included an enhanced discussion of these issues in Section 3.2.6 (Cost and Implementation Issues) of the revised Assessment Document.

Background, matrix effects, and sources of variance (MC-10)

Comment: The Assessment Document gives considerable discussion to the problems that arise from background, matrix effects, and sources of variance. Topics such as instrument maintenance, reliability, and time stability of calibration standards, anion solubility effects, and related topics are also important to implementation of a method. EPA has done a good job addressing these issues, both in the Assessment Document and in Method 1631B.

Response: EPA appreciates the reviewer's positive feedback.

QC charts (MC-11)

Comment: Good QC would include QC charts that identify statistically significant loss of response at the MDL or alternate minimum detection level. The discussion in the Assessment Document does not point out the operational difficulty in applying a method-defined MDL to single-laboratory determinations of a few samples.

The evaluation criteria stated in Chapters 3 and 4 of the Assessment Document do not address adequate measures to estimate increased variability near the limit of detection. Nor do they establish rigorous criteria for data acceptance. In practical laboratory operations, techniques like control charts, maintained over time, would provide reliable measures of variability during actual laboratory operations.

Response: EPA agrees that control charts are instructive in tracking variability in laboratory performance over time. However, EPA believes they have two potential drawbacks: (1) they do not establish an absolute limit within which

analysis must operate, and (2) continued improvement could lead to unusually stringent limits, that eventually will not be met. To compound the problem, EPA found that some regulatory authorities are penalizing laboratories that do not keep QC charts up to date, even though laboratories are operating within the QC acceptance criteria (limits) of a method. In such cases, a concept that is intended to serve as a useful diagnostic tool for the laboratory is having unanticipated negative consequences. That said, however, EPA agrees that control charts, maintained over time, can be instructive in identifying statistically significant losses of analyte responses in the region of interest. EPA has included a discussion of this issue in Section 3.1.5, Measurement Quality over the Life of a Method, of the revised Assessment Document.

However, the issue of "applying a method-defined MDL to single-laboratory determinations of a few samples" is not one that EPA believes can be addressed by the use of control charts. As noted in the Assessment Document, EPA is not promoting the use of the MDL as a reporting limit. EPA believes that the MDL, as a measure of method sensitivity, is an important consideration in the selection of methods to meet the objectives of a given analytical program and provides a potential metric for judging laboratory performance before analyses begin. The application of a method-defined would apply to any limits derived using any of the procedures considered in the Assessment Document.

EPA also notes that there is no *technical* problem in applying an MDL (or other detection limit concept) to single-laboratory determinations of a few samples. The problems are practicality and cost. Because the MDL is usually determined using seven replicates, application to a few samples would significantly increase the cost for the determination. However, nothing in the Assessment Document suggests that the MDL be determined by an individual laboratory at a specific frequency.

Plots of RSD vs concentration (MC-12)

Comment: Data in the Assessment Document and referenced publications cite the loss of precision for ultra-trace determinations near the limit of detection. This effect was plotted in a paper presented at the "22nd International Symposium and shows that relative standard deviation (RSD) decreases rapidly from a high level at low concentrations and reaches an asymptote at high concentrations.

Response: EPA provided plots of RSD vs concentration in Appendix C to the Assessment Document. For nearly all analytes and determinative techniques these plots show the same effect as the plot presented at the 22nd International Symposium. EPA considered the effects of these plots on detection and quantitation concepts in its assessment.

Outliers (MC-13)

Comment: The subject of outliers was given limited attention in Chapter 3 of the Assessment Document. Outlier treatment is a statistically valid area of data treatment. Cochran's test, and single/double Grubbs tests are useful in evaluating interlaboratory data sets to determine outliers and stragglers. Other classical outlier tests could also be evaluated in examining the data sets used in the Assessment Document.

Response: EPA agrees and has included a discussion of outlier treatment in the revised MDL procedure being proposed today.

Prescriptive vs descriptive use of lower limits of measurement (MC-14)

Comment: EPA typically walks a thin line in defining descriptive vs prescriptive procedures. Regulatory requirements built into EPA final rules are very difficult to change and cause a high level of legal liability to laboratories and data users. It is important for EPA to build as much flexibility as possible into CWA methods in order to prevent "locking" unreasonable or unsound procedures into final rule methods.

Response: EPA agrees and has built flexibility in recent methods, such as EPA Method 1631.

National vs local standards of measurement, and NPDES uses (MC-15)

Comment: By law, local restrictions must be as stringent as federal rules. This process is addressed in current law. NPDES permitting is well established in the U.S. There should not be any unusual legal or procedural difficulties that arise from a review of this Assessment Document.

Response: EPA's intent in mentioning local standards of measurement was to determine if local standards should be a consideration in detection and quantitation limit concepts. EPA agrees with the peer reviewer's opinion that there should not be unusual legal or procedural difficulties that arise from EPA's assessment of detection/quantitation limit concepts.

Use of a pair of procedures (MC-16)

Comment: EPA has stated in the draft Assessment Document that one primary procedure is needed for clarity and to avoid confusion among stakeholders. If alternate procedures are needed, the EPA Clean air Act system of reference and equivalent methods has worked well, and could be a model for EPA to follow under the Clean Water Act.

Response: The system of reference methods used under the Clean Air Act (CAA) is similar to the existing "alternate test procedure" (ATP) program for

analytical methods currently used in the Clean Water Act (CWA) and the Safe Drinking Water Act (SDWA). The difference between the ATP program and the use of alternate procedures for establishing detection and quantitation limits, is that in an ATP program, the goal is clear and agreed upon, whereas there remain fundamental theoretical issues surrounding detection and quantitation.

For example, when a test procedure is developed for use in CAA or CWA programs, the reference method is designed to measure Analyte X, in Matrix Y, at some concentration related to a regulatory need (e.g., a permit limit). Alternative procedures may be capable of making measurements of Analyte X in Matrix Y, at the level of concern using different approaches. Thus the demonstration of equivalency between the reference method and a possible alternative method is judged using a metric that consists of Analyte X, Matrix Y, and the level of concern, as well as other aspects of method performance. In contrast, the primary differences between the EPA MDL/ML concepts and potential alternatives, such as the ASTM IDE and IQE, are related to different interpretations as to how detection and quantitation limits should be derived and applied. (These differences are described in the Assessment Document.) Therefore, EPA does not believe that a permutation of existing ATP programs is likely to be an effective model for assessing other detection and quantitation procedures.

However, EPA is willing to consider that an analytical method from a VCSB or other source may be acceptable for approval at 40 CFR part 136 and use in CWA programs even if it employs an alternative procedure for establishing method sensitivity. For example, consider the theoretical situation of an ASTM method for the determination of an analyte regulated under the NPDES program that uses the IDE or IQE to describe method sensitivity *and* for which the value of the IDE or IQE was below the relevant regulatory limit. EPA would evaluate the overall performance of such a method for approval at 40 CFR part 136, despite the fact that the method did not contain an MDL determined using the Appendix B procedure.

In response to this comment, EPA enhanced the discussion of this issue in the revised Assessment Document.

VCSB procedures (MC-17)

Comment: EPA has strived to include voluntary consensus standards body (VCSB) detection and quantitation concepts in the Assessment Document and has conducted an extensive review and discussion of VCSB procedures. International VCSBs, including those from non-governmental organizations (NGOs) should be included.

Response: EPA has included international VCSBs, including NGOs, in its assessment by including detection and quantitation approaches from ASTM-International, International Organization for Standardization (ISO), International Union of Pure and Applied Chemistry (IUPAC), and the American Chemical Society (which is one the largest *international* scientific organizations in the world). However, as discussed in the Assessment Document, ASTM-International is the only one of these organizations that has developed detailed procedures for determining detection and quantitation limits (i.e., the IDE and IQE).

Censoring data and degradation of method performance over time (MC-18)

Comment: Method flexibility is discussed in the Assessment Document and considers time-dependent modifications to rigid methods. EPA is gaining experience in this area and newer methods do address this concern. EPA has developed an evolving method development process that has been shown to be responsive to this issue.

Response: EPA appreciates the positive feedback.

Comments on the Criteria used to Evaluate Detection and Quantitation Limit Concepts (MC-19, MC-20, MC-21, MC-22, MC-23, MC-23b)

Comment: The six criteria should provide a vigorous review of the conditions set out in the draft Assessment Document. *(MC-23b)*

Response: EPA agrees and appreciates the positive feedback.

Comment: Scientific validity is defined in two ways: legal reliability and scientific practice. Criterion 1 is defined by U.S. Supreme Court decisions defining expert testimony. Scientific validity is based on publication in the open literature, competent peer review and general acceptance in the scientific community. The primary detection and quantitation limit procedures evaluated by EPA appear to meet the conditions for legal reliability and scientific practice stated in Criterion 1 of the Assessment Document. One open question concerns "standards." The Assessment Document interprets this condition to mean well-documented methodology. U.S. Constitutional law intended metrology, the legal recognition of reference measures, as a Federal responsibility. If the court's intent was to include legal measures (metrology) as part of expert testimonial evidence, the need for a defined reference material, or EPA audit standard, is implied and should be considered. This reviewer is not competent to answer this legal question. All the other elements of Criterion 1 seem to be addressed in the Assessment Document. *(MC-19)*

Response: EPA appreciates the feedback. EPA wishes to emphasize that EPA used the conditions established in the Daubert decision as a *guideline* for establishing

scientific validity because the scientific community has not established a consensus definition of term. EPA believes that some flexibility in applying these conditions to an evaluation of scientific validity is reasonable.

Comment: Criterion 2 (demonstrated method performance) appears to be met under EPA Method 1613B and other EPA-cited methods used as examples in the Assessment Document. Measurement of variability and defined method expectations may require a special study that addresses all candidate alternate procedures and parameters that interested stakeholders deem significant. *(MC-20)*

Response: EPA agrees that the performance of any method developed for use in CWA programs should be studied to establish method performance characteristics (including sensitivity). In the Assessment Document, EPA reaffirms use of the MDL and ML to establish sensitivity, but as noted in the response to a previous comment, EPA also is willing to approve methods submitted by other organizations that rely on alternate sensitivity approaches. In such cases, EPA would require that method performance be characterized and that the sensitivity limits established in the subject method be at or below the applicable regulatory limits.

Comment: Criterion 3 addresses performance of a procedure that is practical and affordable for use by a single laboratory. This criterion is important because it isolates theoretical estimators and large demonstration studies (interlaboratory and intralaboratory comparisons) from the fundamental application of any EPA method for single or small numbers of determinations. The most common situation is a laboratory performing many types of analysis but must perform EPA Method 1631 on a periodic basis where reproducibility is poor.

Criterion 3 should judge method ruggedness and appropriate quality control to make a method reliable and well as "laboratory friendly." Criterion 3 also addresses cost which is very important, but this may need to be a final estimator after other parameters are settled. Criterion 3 should be strengthened both in performance discussions and proposed method modifications. *(MC-21)*

Response: EPA agrees. A large number of small laboratories operate in the U.S., and these small laboratories do not have the capability of conducting interlaboratory studies to establish detection and quantitation limits. Therefore, detection and quantitation procedures should be practical and affordable for use by a single laboratory. In response to this reviewer's suggestion, EPA has revised the discussion of Criterion 3 in the Assessment Document (Section 4.3) to address the need for a "laboratory friendly" procedure and enhance the discussion of method ruggedness and appropriate control, particularly as it relates to demonstrating method performance in real world matrices.

Comment: Criteria 4 and 5 address the primary subject matter of the Assessment Document: detectability (assure 99% detection confidence in an experienced laboratory) and quantifiability (assure reliable quantification limit in an experienced laboratory). As such, they are significant and should be maintained. *(MC-22)*

Response: EPA appreciates the positive feedback.

Comment: Criterion 6 addresses conditions in the method that meet federal limits and allow for more stringent application by local regulatory bodies. This criterion is essential and cannot be changed. *(MC-23)*

Response: EPA agrees.

Conceptual soundness of the MDL and ML (MC-24)

Comment: The MDL and ML have stood the test of time and provide a proven methodology which meets defined evaluation criteria stated in the Assessment Document.

Response: EPA agrees and appreciates the positive feedback.

EPA's assessment appears to be valid (MC-25)

Comment: The assessment in Chapter 5 of the Assessment Document appears valid based on the stated criteria. Detection using the MDL, in my opinion, is valid. Quantification concepts are subject to a higher degree of scientific challenge and interpretation.

Response: EPA appreciates the positive feedback and agrees that the use of quantitation limits are likely to be subject to a higher degree of challenge and interpretation. For example, another peer reviewer noted that nearly all quantitation limit approaches have nothing do with whether the measurements are actual quantitative, and that "the only real criterion for a quantitation limit is that the instrument generate a recognizable signal." EPA has revised the Assessment Document (see Section 4.4) to include an enhanced discussion of this issue, and in today's proposed rule is soliciting comments on whether a quantitation limit is necessary and, if yes, whether the ML is an appropriate quantitation limit.

Evaluation of alternate procedures (MC-26)

Comment: The review process might be strengthened if EPA were to suggest experiments to evaluate alternate detection-quantitation procedures. To completely test the six criteria stated in the Assessment Document, a tailored validation study would need to be designed and performed.

Response: An objective of EPA's variability versus concentration studies described in Sections 1.3.2.1 - 1.3.2.3 of the Assessment Document was to resolve the detection/quantitation issue. The study covered some of the most commonly used techniques for environmental measurement and a range of concentrations from 0.1 to 100 times the MDL. The data showed that detection and quantitation limits can be variable and greatly dependent on the data set used. Based on results from these and other studies, EPA concluded that judgment is required in setting detection and quantitation limits, regardless of how much data are collected. Therefore, EPA does not believe that additional experiments would resolve the issue.

Agreement with EPA's conclusions in Chapter 6 of the Assessment Document (MC-27)

Comment: I agree with EPA's conclusions, based on the conditions laid out in Chapters 3, 4, and 5. Furthermore, EPA has documented that the MDL is a sound estimator of initial signal response in a broad range of analytical methods. The MDL has stood the test of time and I could not find a convincing statistical argument to replace the MDL. However, alternate methods do demonstrate potential improvements to the MDL implementation (i.e., criteria for initial spike determination and selection). The ML and other candidate procedures for quantitation limits show significant variability.

Response: EPA appreciates the support for the MDL and is proposing further improvements to the MDL, such as additional criteria for initial spike determination and selection.

Maintaining the MDL and ML (MC-28)

Comment: The overall conclusion from reading the Assessment Document is that EPA has made a strong case for maintaining the MDL and ML as reference procedures. Most of the alternate detection and quantitation procedures evaluated in the Assessment Document are rejected because they have not been tested extensively in the manner that EPA challenges its internal procedures before publication for regulatory applications. Candidate alternate procedures were drafted by non-governmental organizations (NGOs) as generally applicable without consideration for the legal constraints placed on EPA. EPA procedures are formed around legally defined analyte lists (e.g., the Priority Pollutant List), producing limited numbers of analytes and "bright-line" legal limits that define compliance vs violation. NGOs usually do not create method criteria based on these legal constraints. This disconnect, seen between EPA and candidate alternate procedures, is to be expected. EPA has handled this problem in other media (e.g., the Clean Air Act) by establishing one or more EPA reference procedures, then establishing minimum criteria for equivalency. This could be done with candidate alternate procedures if they contain statistically sound principles that allow equivalent performance.

Response: EPA thanks the reviewer for acknowledging that EPA has made a strong case for maintaining the MDL and ML as reference procedures. As explained in a previous response, the acceptance of alternate detection and quantitation limit approaches cannot be exactly compared to the procedures for evaluating and accepting alternate test procedures under the Clean Air Act or Clean Water Act. EPA believes that rather than establishing specific criteria for demonstrating the validity of alternative sensitivity approaches, it may be better to simply evaluate methods that use alternative approaches for establishing sensitivity. A test method would be accepted if the sensitivity (and overall performance) of the method has been characterized and is sufficient to meet EPA's regulatory needs.

No need to suspend EPA regulatory programs (MC-29)

Comment: There are no strong statements in the Assessment Document that would cause a level of concern needed to suspend EPA regulatory programs or methodology pending additional review.

Response: EPA has not suspended rulemaking pending resolution of the detection/ quantitation issue.

Variability of low-level data (MC-30)

Comment: Data supplied with the Assessment Document show that low-level samples are subject to higher relative variance and should be treated differently from data above an agreed quantitation limit.

Response: EPA has suggested that data below the quantitation limit should not be used for setting permit limits or other regulatory purposes in EPA's CWA programs. However, EPA recognizes the authority of the States to use the detection limit or other threshold when necessary to protect human health or the environment.

Aware of no further data (MC-31)

Comment: I am not aware of any specific data sets that could elucidate the various approaches and challenges listed in the Assessment Document. Even if such databases exist, it would be very difficult to make the appropriate computations and solicit adequate reviews from interested parties given the limitations of the six evaluation criteria.

Response: Because the reviewer did not state the limitations of the six evaluation criteria, we could not determine the potential effect of the comment on existing or postulated databases.

Interlaboratory issues (MC-32)

Comment: EPA has done an adequate job showing performance of EPA methods, especially defining detection (e.g, MDL). EPA has presented extensive data on interlaboratory studies that demonstrate method performance for a number of EPA regulatory procedures.

Response: EPA thanks the reviewer for the acknowledgment.

EPA should consider the need for a tailored demonstration study (MC-33)

Comment: To fully evaluate alternate approaches, a cooperative study should be performed that is designed with input from all settlement participants, and interested outside laboratory professionals. That study could include spike levels, blank and zero determination, intralaboratory variability, ruggedness testing, "pairs" determinations, use of "real world" samples, evaluation of outlier criteria, sufficient replicates to challenge statistical models, and reproducibility versus repeatability (*e.g*., single unbroken series of determinations, versus, series performed on different dates after set up and calibration).

Response: As noted in a response to another reviewer, EPA's objective in conducting the Multi-Technique Variability Study (the Episode 6000 study), was to create a database that would help answer questions concerning detection and quantitation limits. The study design reflected comments and concerns raised by interested and concerned parties (e.g., settlement participants, laboratory scientists, statisticians, and regulatory personnel) at various meetings and conferences. The study was conducted using 7 replicates at 16 concentration levels and 11 of the most commonly used analytical techniques. EPA considered conducting the study at multiple laboratories to evaluate interlaboratory variability, but the additional factor of 6 to 10 in cost made this approach cost-prohibitive. EPA believes, however, that the Episode 6000 study shows that detection limits can be variable and that some judgement will always be required in evaluating detection limit determinations, regardless of how much data is collected.

EPA also notes that the Agency has involved all interested and concerned parties in development of all recent wastewater methods. For example, the Electric Power Research Institute (EPRI) provided funding for spikes at additional concentrations in EPA's interlaboratory validation of EPA Method 1631 and EPA involved the international community in the interlaboratory validation of EPA Method 1613.

EPA should consider further validation of EPA Method 1631 (MC-34)

Comment: Since the subject legal settlement specifically addressed mercury using EPA Method 1631B, the focus of any collaborative study to answer questions

raised in the Assessment Document and legal challenges, should include this specific method. In a joint validation study, it would be useful if EPA incorporated uniform procedures to be followed for any alternate procedures that supplant EPA numbered methods.

Response: EPA asserts that Method 1631 is sufficiently validated, Revision E to Method 1631 was published in a final rule on October 29, 2002 (67 FR 65876) to comply with requirements of the Settlement Agreement. Results from the interlaboratory validation of Method 16311 were used in the Agency's evaluation of detection and quantitation limits, as described in the revised Assessment Document.

EPA should consider better method equivalency and method flexibility (MC-35)

Comment: Simple equivalency procedures are routinely specified by several EPA Offices. For air determinations, EPA provides a generic method protocol to demonstrate method performance. This is used to show that an alternate procedure is suitable for reporting accurate data for regulatory purposes. For stack methods, four (4) concurrent determinations in the same source are required. EPA could use this type of process to set and demonstrate simplified equivalency criteria for existing EPA water and waste water methods. Such guidance (method equivalency and flexibility) will become more critical as detection limits are driven lower, additional analytes are required, and more complex matrices are added to areas of regulatory concern.

Response: EPA thanks the reviewer for the comment, but notes that analytical test method equivalency and flexibility are beyond the scope of this assessment, which is focused solely at approaches for establishing analytical detection and quantitation limits.

EPA should consider the need for an approved reference material and audit standard (MC-36)

Comment: Interlaboratory performance is highly variable using modern ultra-trace methods like EPA Method 1631. When small batches of samples arrive at most environmental laboratories, they are scheduled in series with other methods and different analytes than mercury. These samples must be checked in, records verified, and proper storage and chain-of-custody implemented. At that point the appropriate equipment must be started and calibrated. This process is the worst case, intermittent analyses where all the causes of variability, and sensitivity loss, are maximized. This means quality control on every batch of samples becomes critically important. EPA validation studies, which demonstrate the optimum method performance, are useful guidance; however, one-time optimum performance does not reflect batch-to-batch data quality in actual operation. In the real

world non-optimum operation is the rule, not the exception. This problem is exacerbated with ultra-trace methods.

Response: EPA thanks the reviewer for this comment and will consider this issue during future method development and revision efforts. EPA also wishes to note that Revision E to EPA Method 1631, promulgated on October 29, 2002 (67 FR 65876), contains more extensive batch quality control (QC) than any EPA wastewater method promulgated to date. This QC is mandatory.

Note: In addition to the comments detailed above, this peer reviewer also provided a suite of detailed comments and suggestions concerning EPA Method 1631B. EPA thanks the reviewer for these detailed comments, but notes that specific analytical test methods are beyond the scope of this assessment, which is focused solely at approaches for establishing analytical detection and quantitation limits. EPA will retain the reviewer's comments on Method 1631 for further consideration during any future revision of that test method.